Fast, Marginalised Bayesian Transient Searching

Sam Leeney

Handley Lab, University of Cambridge

Collaborators and discussion partners across SKA/CHIME

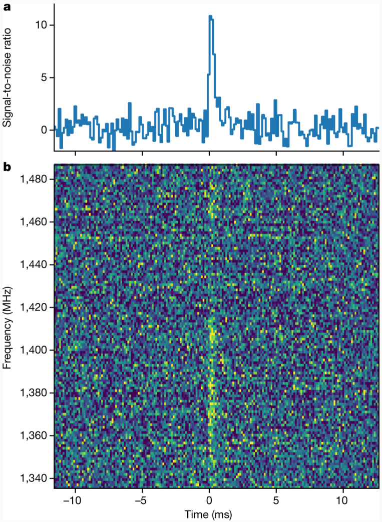
Key message

Searching directly on raw time–frequency data with a marginalised Bayesian model lets us keep information, model anomalies jointly, and stay physics-aware.

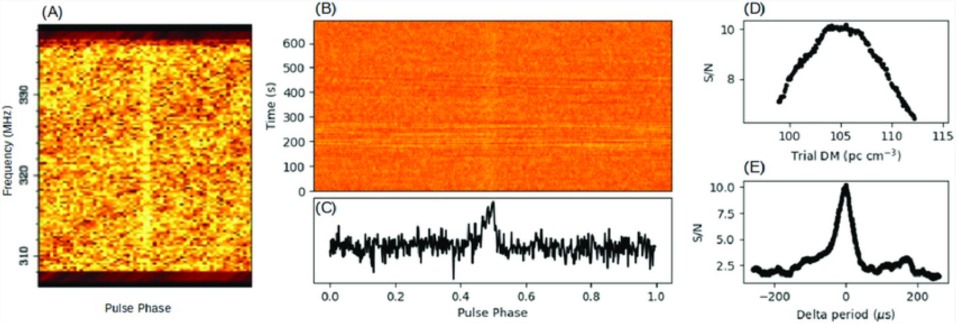
- Compression loses power; search on raw cubes to avoid throwing signals away.
- Fit and flag together: anomaly probability lives inside the likelihood, not after the fact.
- Use physics-informed priors (DM, width, arrival time) to steer the search space.
- The math parallelises; GPU acceleration makes this practical.

What are radio transients?

Fast radio bursts

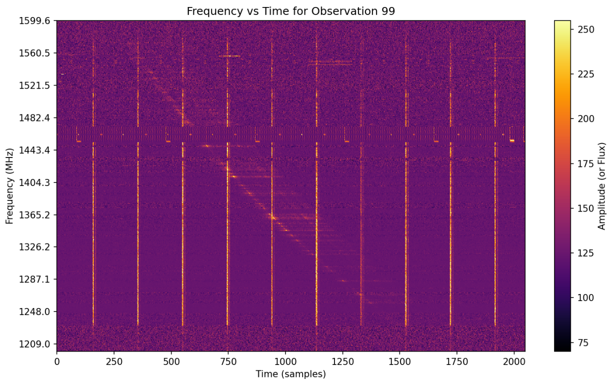


Pulsars and repeating bursts



What do they look like in data?

Dynamic spectrum before dedispersion

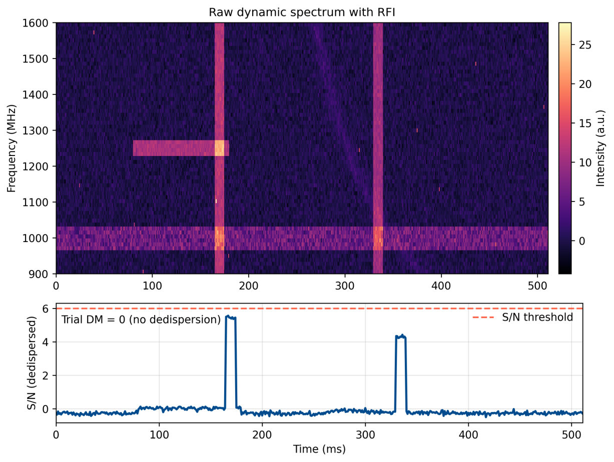


- Typically we see transients like in the previous images.
- At the telescope they look like this: dispersed across frequency.
- Also mixed with interference.
- Typical pipelines compress this data before fitting; we want to operate in this space entirely.

Standard search algorithms

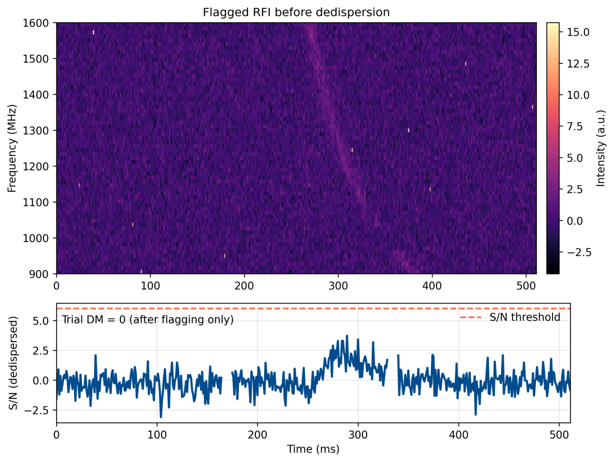
How do standard search algorithms work?

Next: RFI flagging, dedispersion sweeps, and S/N thresholding in practice



What you see

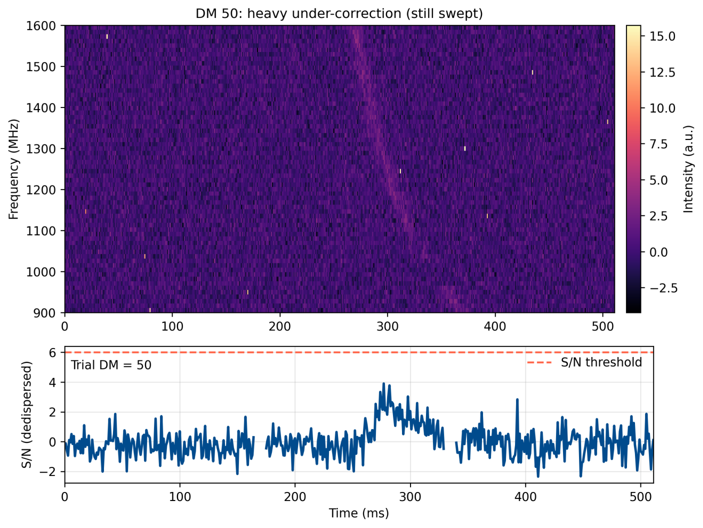
- Dispersed sky pulse buried under thermal noise.
- Broadband glitch and narrowband line RFI dominate the collapsed S/N.
- No dedispersion yet: pulse power is smeared out in time.



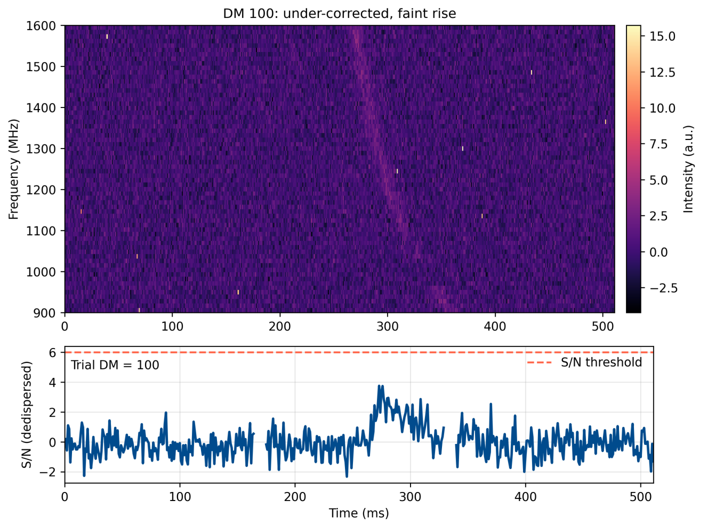
Effect on the data

- Mask out the obvious broadband burst and persistent narrowband line.
- Residual dynamic spectrum still contains the dispersed pulse + noise.
- Collapsed S/N improves slightly but is still below threshold.

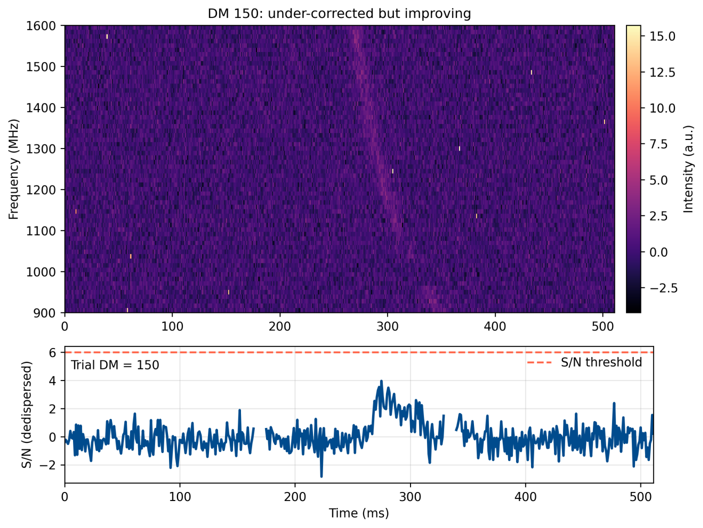
Step 2: dedispersion sweep (low DM)



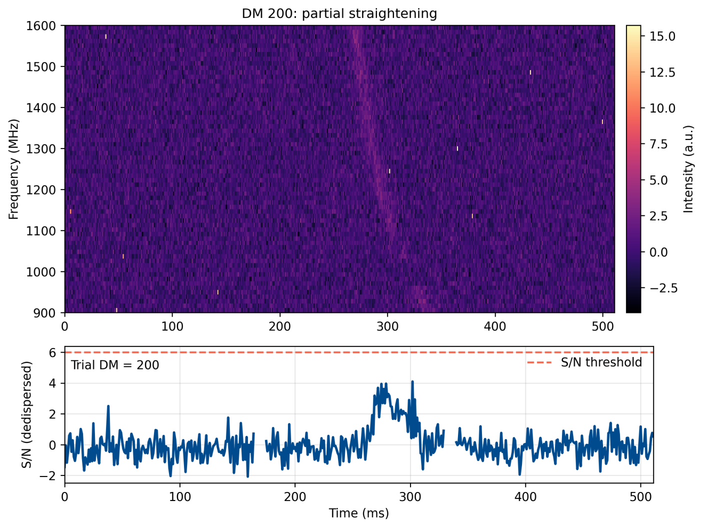
Dedispersion sweep (DM 100)



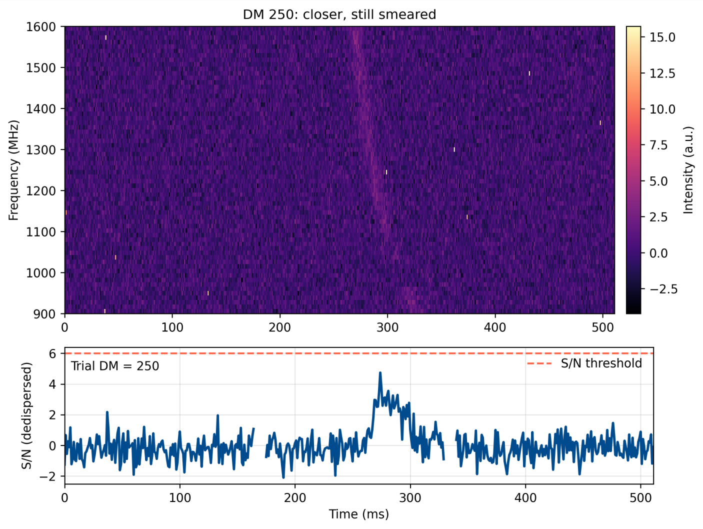
Dedispersion sweep (still low DM)



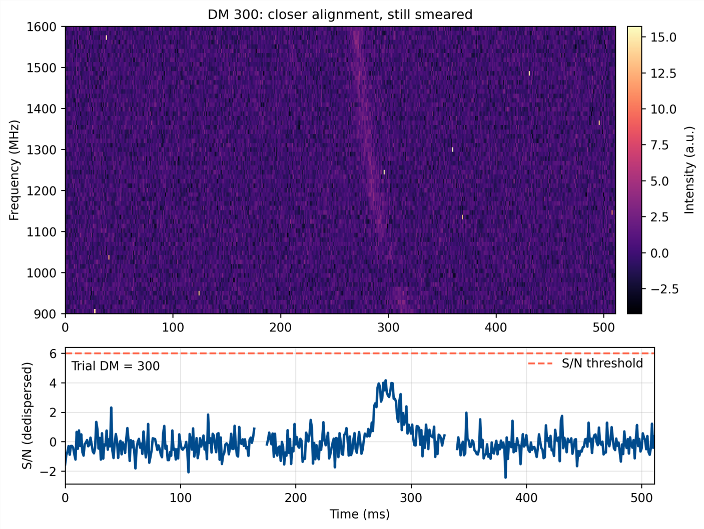
Dedispersion sweep (DM 200)



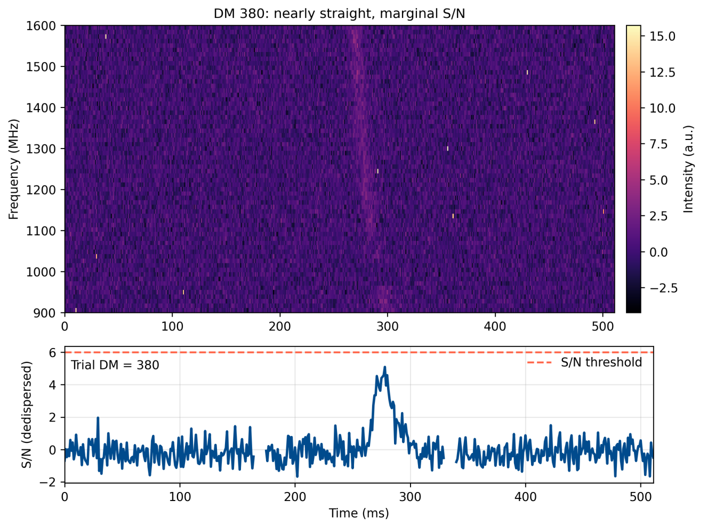
Dedispersion sweep (DM 250)



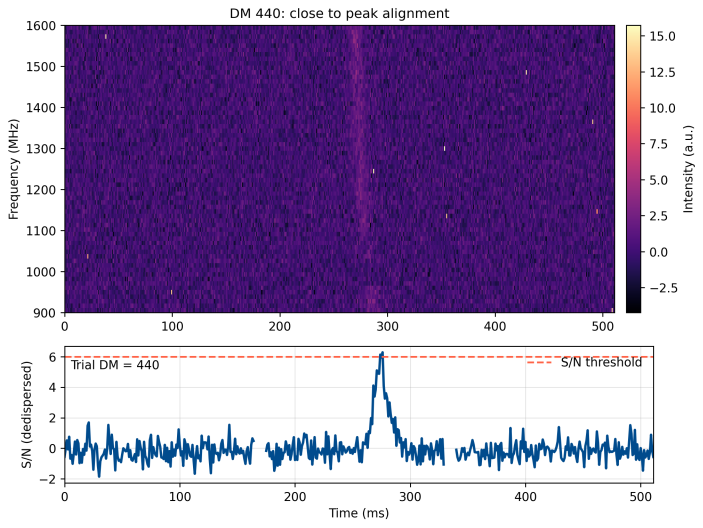
Dedispersion sweep (closer DM)



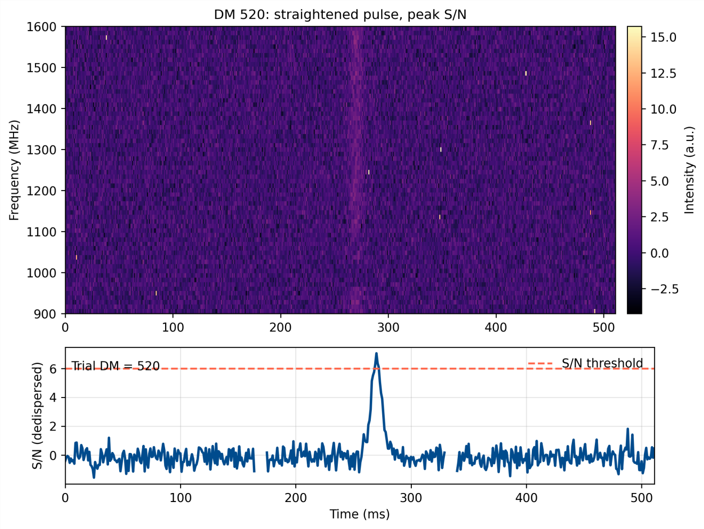
Dedispersion sweep (DM 380)



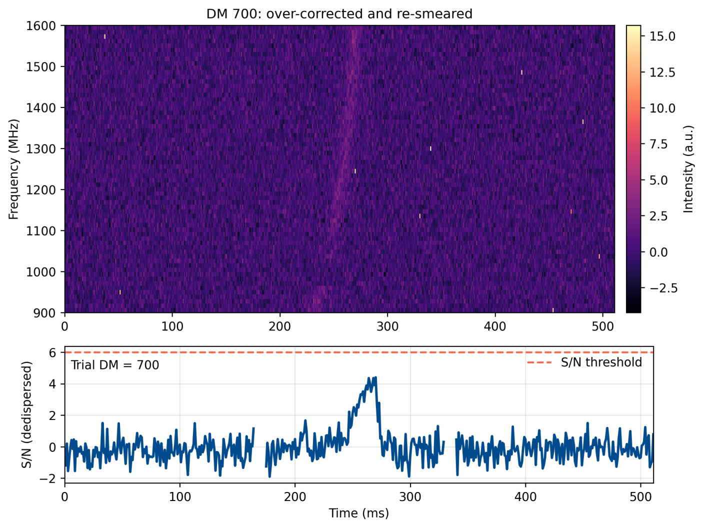
Dedispersion sweep (DM 440)



Dedispersion sweep (approaching DM)



Dedispersion sweep (over-corrected)



Bayesian alternative

Idea: keep the raw information

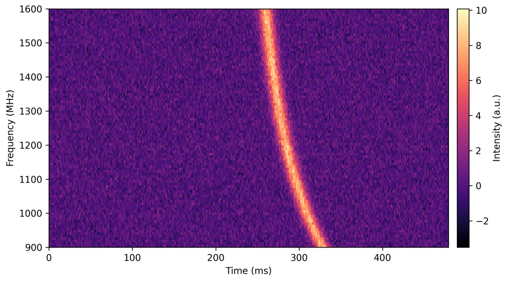
- Choose search parameters from physics-informed priors (DM, width, arrival window).
- Fit and flag simultaneously so no information is discarded to pre-filtering.
- Stay in the native time–frequency cube—no irreversible compression first.
- Make it GPU-fast by keeping the marginalisations analytic and parallel.

Step 1: generative model

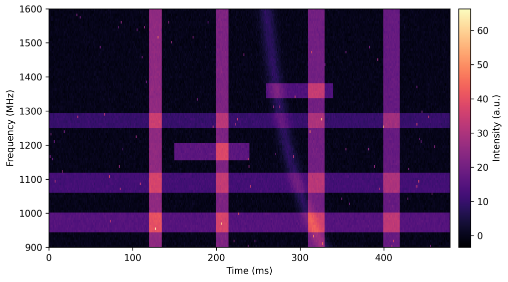
$$I_{ij} = \frac{A}{\sqrt{2\pi}\,\sigma'_{\mathrm{frb}}} \exp\left(-\frac{\left[t_j - t'_0 - K\,\mathrm{DM}'/f_i^2\right]^2}{2\,{\sigma'_{\mathrm{frb}}}^2}\right) + \epsilon_{ij},$$

- Likelihood: $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{\mathrm{noise}})$; evaluate on the full cube.
- Sample with nested sampling to obtain posteriors and evidence for model comparison.

Step 1: simple vs. RFI-contaminated pulse



Clean simulated pulse



Same pulse with strong RFI (not yet modelled)

Bayesian anomaly detection

Bayesian anomaly detection (Bernoulli mask)

- Each pixel carries a latent mask $\epsilon_{ij} \in \{0, 1\}$: $\epsilon = 1$ marks an anomaly (RFI), $\epsilon = 0$ is nominal.
- Prior on the mask: $P(\epsilon_{ij}) = p^{\epsilon_{ij}}(1 - p)^{1-\epsilon_{ij}}$ with p as the anomaly rate.
- Full likelihood (before marginalisation):

$$P(\mathbf{I}, \epsilon \mid \theta, p) = \prod_{ij} \left[(1 - p)L_{ij}(\theta)\right]^{1-\epsilon_{ij}} \left(\frac{p}{\Delta}\right)^{\epsilon_{ij}},$$

where $L_{ij}(\theta)$ is the nominal Gaussian likelihood and Δ is a broad anomaly scale.

Marginalising anomalies without 2^N masks

- Exact marginal in log-space (only expression we keep):

$$\log P(\mathbf{I} \mid \theta, p) = \sum_{ij} \log\Big((1-p)L_{ij}(\theta) + \frac{p}{\Delta}\Big)\,.$$

- Expanding shows the Occam term that downweights extra mask freedom:

$$\log P = \underbrace{\sum_{ij} \log\big((1-p)L_{ij}\big)}_{\text{fit term}} + \underbrace{\sum_{ij} \log\left(1 + \frac{p/\Delta}{(1-p)L_{ij}}\right)}_{\text{Occam penalty}}\,.$$

- Maximising over ϵ for intuition gives

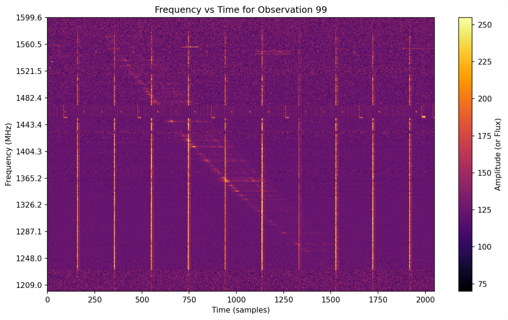
$$\log P(\mathbf{I} \mid \theta, \epsilon^{\max}) = \sum_{ij} \log\Big(\max\big((1-p)L_{ij},\, p/\Delta\big)\Big),$$

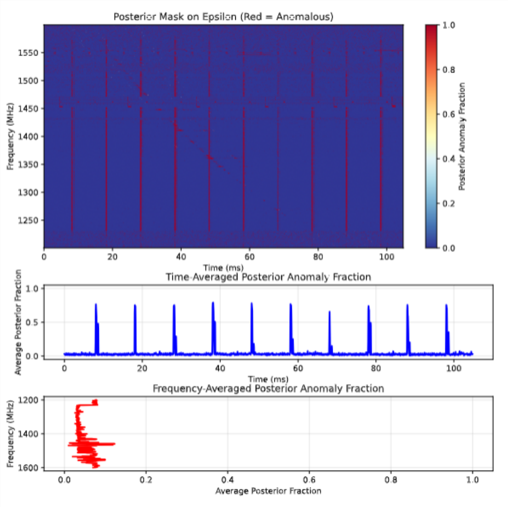
  and the marginal log-evidence sits below this by the penalty term above.

- Result: a tractable likelihood that jointly fits the pulse and flags RFI probabilistically, yielding evidence $\mathcal{Z}_{\text{signal+anomaly}}$ for detection.

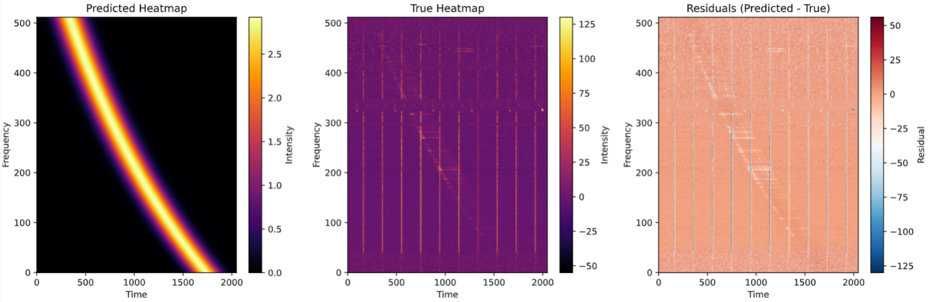
Real data fits

Run on real data (flagged anomalies)





Run on real data (fit)



Flagging detections

- We flag when the Bayes factor exceeds 1:

$$\mathcal{B} = \frac{\mathcal{Z}_{\text{signal+anomaly}}}{\mathcal{Z}_{\text{noise-only}}} > 1.$$

- Posterior on p shows how much of the cube was treated as interference.

Problem... too slow

- Even with JAX/Numpyro-style acceleration, full inference takes minutes per candidate.
- At common resolutions, evaluating the Bayes factor can take minutes per ms of data.
- We need an analytic marginalisation for the evidence to keep up with survey rates.

Analytic Bayes factor

Goal: closed-form Bayes factor

$$K = \frac{Z_1}{Z_0}, \qquad Z_m = \int p(I \mid \theta, M_m)\, p(\theta \mid M_m)\, d\theta$$

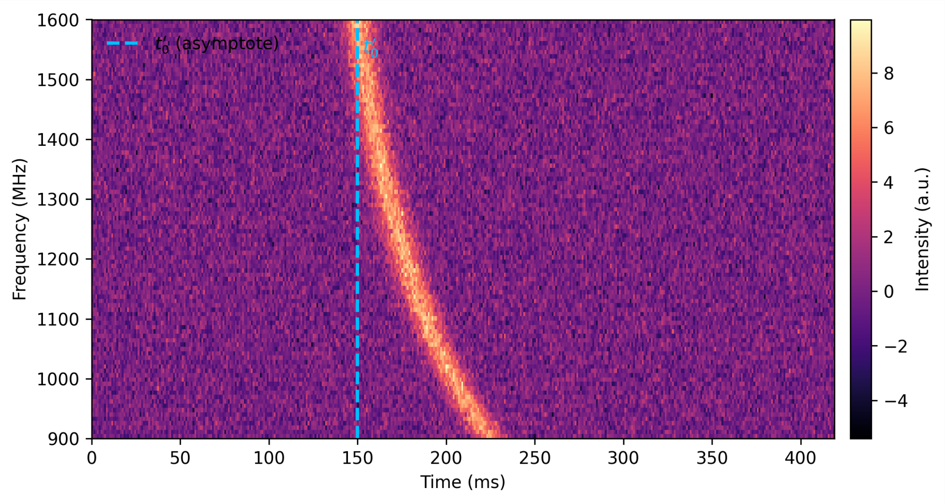
- Treat all template parameters as nuisance; integrate them out analytically where possible.
- Keep only a small numerical search over $(\mathrm{DM}', \sigma'_{\mathrm{frb}})$ and discrete t'_0.
- Result: fast, calibrated evidence for detection.

Signal model (from Step 1)

$$I_{ij} = \frac{A}{\sqrt{2\pi}\,\sigma'_{\mathrm{frb}}} \exp\left(-\frac{\left[t_j - t'_0 - K\,\mathrm{DM}'/f_i^2\right]^2}{2\,{\sigma'_{\mathrm{frb}}}^2}\right) + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

- Parameters to marginalise: t'_0 (discrete), amplitude A, noise σ, track shape $(\mathrm{DM}', \sigma'_{\mathrm{frb}})$.

Visualising t_0' (arrival asymptote)



Marginalising t_0' (discrete time index)

$$I_{ij} = \frac{A}{\sqrt{2\pi}\,\sigma'_{\mathrm{frb}}} \exp\left(-\frac{\left[t_j - t_0' - K\,\mathrm{DM}'/f_i^2\right]^2}{2\,{\sigma'_{\mathrm{frb}}}^2} \right) + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

- Treat t_0' as a discrete prior over time bins (often uniform).
- Condition on t_0' and evaluate in parallel across time bins:

$$Z_1(t_0') = \iint p(I \mid t_0', A, \sigma, \mathrm{DM}', \sigma'_{\mathrm{frb}})\, p(A, \sigma, \mathrm{DM}', \sigma'_{\mathrm{frb}})\, dA\, d\sigma\, d\phi$$

with $\phi = (\mathrm{DM}', \sigma'_{\mathrm{frb}})$.

- Full evidence: $Z_1 = \sum_{t_0'} p(t_0')\, Z_1(t_0')$; embarrassingly parallel over t_0'.

Marginalising amplitude A

$$I_{ij} = \frac{A}{\sqrt{2\pi}\,\sigma'_{\mathrm{frb}}} \exp\left(-\frac{\left[t_j - t'_0 - K\,\mathrm{DM}'/f_i^2\right]^2}{2\,{\sigma'_{\mathrm{frb}}}^2}\right) + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

- Write $\mu_{ij} = A\,\psi_{ij}(t'_0, \mathrm{DM}', \sigma'_{\mathrm{frb}})$.
- Gaussian prior on $A \Rightarrow$ complete the square and integrate:

$$A \mid \psi, I, \sigma^2 \sim \mathcal{N}(\hat{A}, \Lambda_A^{-1}), \quad \hat{A} = \frac{\psi^\top I}{\psi^\top \psi + \sigma_A^{-2}\sigma^2}.$$

- Plugging back yields a marginal likelihood depending only on σ^2 and ϕ.

Marginalising noise σ^2

$$I_{ij} = \frac{A}{\sqrt{2\pi}\,\sigma'_{\mathrm{frb}}} \exp\left(-\frac{\left[t_j - t'_0 - K\,\mathrm{DM}'/f_i^2\right]^2}{2\,{\sigma'_{\mathrm{frb}}}^2}\right) + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

- Prior $\sigma^2 \sim \mathrm{Inv}\text{-}\Gamma(\alpha_0, \beta_0)$, likelihood Gaussian on residuals.
- After integrating out A, residual sum of squares S gives:

$$p(I \mid t'_0, \phi) \propto \left(\beta_0 + \tfrac{1}{2} S\right)^{-(\alpha_0 + N_{\mathrm{pix}}/2)}.$$

- Noise variance disappears analytically; no sampler needed.

Laplace for $(\mathrm{DM}', \sigma'_{\mathrm{frb}})$

$$I_{ij} = \frac{A}{\sqrt{2\pi}\,\sigma'_{\mathrm{frb}}} \exp\left(-\frac{\left[t_j - t'_0 - K\,\mathrm{DM}'/f_i^2\right]^2}{2\,\sigma'_{\mathrm{frb}}{}^2} \right) + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

- Remaining continuous track parameters $\phi = (\mathrm{DM}', \sigma'_{\mathrm{frb}})$.
- Laplace about $\hat{\phi}$ with Hessian H:

$$\log Z_1(t'_0) \approx \ell(\hat{\phi}) + \frac{d}{2}\log(2\pi) - \tfrac{1}{2}\log|H|, \quad d = 2.$$

- Sum over t'_0 and form $K = Z_1/Z_0$ for a fast Bayes factor per candidate.
- Cuts evidence time from minutes to ~10s of ms per candidate; still need to fold in RFI handling.

Speed-up from analytic evidence